

On the Inclusion Relation of Reproducing Kernel Hilbert Spaces

Haizhang Zhang^{*} and Liang Zhao[†]

Abstract

To help understand various reproducing kernels used in applied sciences, we investigate the inclusion relation of two reproducing kernel Hilbert spaces. Characterizations in terms of feature maps of the corresponding reproducing kernels are established. A full table of inclusion relations among widely-used translation invariant kernels is given. Concrete examples for Hilbert-Schmidt kernels are presented as well. We also discuss the preservation of such a relation under various operations of reproducing kernels. Finally, we briefly discuss the special inclusion with a norm equivalence.

Keywords: inclusion, embedding, refinement, reproducing kernels, reproducing kernel Hilbert spaces

1 Introduction

Reproducing kernel Hilbert spaces (RKHS) are Hilbert spaces of functions on which point evaluations are always continuous linear functionals. They are the natural choice of background spaces for many applications. First of all, thanks to the existence of an inner product, Hilbert spaces are the normed vector spaces that are well-understood and can be handled best. Secondly, the inputs for many application-oriented algorithms are usually modeled as the sample data of some desirable but unknown function. Requiring the sampling process to be stable seems to be a necessity. Mathematically, this is synonymous with desiring point evaluation functionals to be bounded. For these reasons, RKHS are widely applicable in probability and statistics [2, 19], dimension reduction [9], numerical study of differential equations [5, 10], generalizations of the Shannon sampling theory [13, 26], and approximation from scattered data [22]. Moreover, an RKHS possesses a unique function, named a reproducing kernel, which represents point evaluations on the space. Reproducing kernels are able to measure the similarity between inputs and could save the calculation of inner products in a feature space [17]. This gives birth to the “kernel trick” in machine learning and makes RKHS the popular underlying feature spaces for applications in the field. As a result, reproducing kernel based methods are dominant in machine learning [4, 7, 17, 18, 21].

Despite the wide applications of RKHS, there are some important theoretical issues that are not well-understood. This paper is devoted to the inclusion relation between RKHS, that is, given two

^{*}School of Mathematics and Computational Science and Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou 510275, P. R. China. E-mail address: zhhaizh2@sysu.edu.cn. Supported in part by Guangdong Provincial Government of China through the “Computational Science Innovative Research Team” program.

[†]Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA. E-mail address: lzhao04@syr.edu. Supported in part by US Air Force Office of Scientific Research under grant FA9550-09-1-0511.

reproducing kernels, we are interested in whether the RKHS of one reproducing kernel is contained by the RKHS of the other. The clarification of this problem is helpful to understand the structure of RKHS and hence is contributive to the theory of reproducing kernels [1]. For instance, the relation is needed in building a multi-resolution decomposition of RKHS. Besides, the study could provide guidelines to the choice of reproducing kernels in machine learning. There are many reproducing kernels in the literature. In a particular application, the selection of reproducing kernels is usually critical to the success of a learning algorithm. While there are no well-recognized guidelines in making such a decision, avoiding overfitting or underfitting is usually the first principle. When overfitting or underfitting occurs, a remedy is to change the current reproducing kernel so that the RKHS of the new kernel becomes smaller or larger compared to that of the existing kernel. Understanding the inclusion relation between RKHS could help achieve such an update of reproducing kernels.

Three characterizations of the inclusion relation of RKHS were established before 1970s [1, 6, 25]. With the advent of machine learning in 1990s, there has been increasing interest in reproducing kernels and RKHS. Many concrete reproducing kernels have emerged in the literature and in applications. Most of them can be conveniently represented by a feature map, which was unknown in the past studies [1, 6, 25]. The purpose of this paper is to provide a systematic study of the inclusion relation of RKHS with focus on the concrete examples of RKHS appeared in machine learning. Recent references [23, 24] studied the embedding relation of RKHS, that is, an equal norm requirement is imposed. As shown by the examples therein, the requirement that two RKHS share the same norm on the smaller space might be demanding and rules out many commonly-used RKHS. For example, the RKHS of a Gaussian kernel can not be properly embedded into the RKHS of another translation invariant reproducing kernel of a continuous type. By relaxing the requirement, we shall see more applications and have more structural results.

The outline of the paper is as follows. We shall discuss characterizations of the inclusion relation in the next section. Sections 3 and 4 are devoted to the investigation of concrete translation invariant and Hilbert-Schmidt reproducing kernels, respectively. Particularly, we shall establish a full table of inclusion relations among popular translation invariant reproducing kernels in Section 3. In Section 5, we discuss the preservation of the relation under various operations of reproducing kernels. In the last section, we shall briefly discuss the special inclusion relation where a norm equivalence is required.

2 Characterizations

We start with introducing some basics of the theory of reproducing kernels [1]. Let X be a prescribed set, which is often referred to as an input space in machine learning. A *reproducing kernel* (or kernel for short) K on X is a function from $X \times X$ to \mathbb{C} such that for all finite pairwise distinct inputs $\mathbf{x} := \{x_j : j \in \mathbb{N}_n\} \subseteq X$, the kernel matrix

$$K[\mathbf{x}] := [K(x_j, x_k) : j, k \in \mathbb{N}_n]$$

is hermitian and positive semi-definite. Here, for the simplicity of enumerating with finite sets, we denote for each $n \in \mathbb{N}$ by $\mathbb{N}_n := \{1, 2, \dots, n\}$. A reproducing kernel K on X corresponds to a unique RKHS, denoted by \mathcal{H}_K , such that $K(x, \cdot) \in \mathcal{H}_K$ for all $x \in X$ and

$$f(x) = (f, K(x, \cdot))_{\mathcal{H}_K} \quad \text{for all } f \in \mathcal{H}_K, x \in X, \quad (2.1)$$

where $(\cdot, \cdot)_{\mathcal{H}_K}$ denotes the inner product on \mathcal{H}_K . There is a characterization of reproducing kernels in terms of feature maps. A *feature map* for a kernel K on X is a mapping from X to another Hilbert

space \mathcal{W} such that

$$K(x, y) = (\Phi(x), \Phi(y))_{\mathcal{W}}, \quad x, y \in X. \quad (2.2)$$

The space \mathcal{W} is call a *feature space* for kernel K . One observes from (2.1) that

$$K(x, y) = (K(x, \cdot), K(y, \cdot))_{\mathcal{H}_K}, \quad x, y \in X.$$

Thus, $\Phi(x) := K(x, \cdot)$, $x \in X$ and $\mathcal{W} := \mathcal{H}_K$ is a pair of feature map and feature space for K . The RKHS of a reproducing kernel can be easily identified once a feature map representation is available. The following result is well-known in machine learning community [14, 17, 23].

For a feature map $\Phi : X \rightarrow \mathcal{W}$, we shall denote by P_Φ the orthogonal projection from \mathcal{W} onto the linear span $\text{span } \Phi(X)$ of $\Phi(X)$.

Lemma 2.1 *If K is a kernel on X given by (2.2) by a feature map Φ from X to \mathcal{W} then $\mathcal{H}_K = \{(\Phi(\cdot), u)_{\mathcal{W}} : u \in \mathcal{W}\}$ with the inner product*

$$(u, \Phi(\cdot))_{\mathcal{W}}, (v, \Phi(\cdot))_{\mathcal{W}}_{\mathcal{H}_K} = (P_\Phi u, P_\Phi v)_{\mathcal{W}}, \quad u, v \in \mathcal{W}. \quad (2.3)$$

In particular, if $\text{span } \Phi(X)$ is dense in \mathcal{W} then \mathcal{H}_K is isometrically isomorphic to \mathcal{W} through the linear mapping $(u, \Phi(\cdot))_{\mathcal{W}} \rightarrow u$.

As an example, we look at the sinc kernel

$$K(x, y) = \text{sinc}(x - y) := \prod_{j=1}^d \frac{\sin \pi(x_j - y_j)}{\pi(x_j - y_j)}, \quad x, y \in \mathbb{R}^d.$$

It can be represented as the Fourier transform of $\frac{1}{\sqrt{2\pi}^d} \chi_{[-\pi, \pi]^d}$, where χ_A is the characteristic function of a subset $A \subseteq \mathbb{R}^d$. In this paper, we adopt the following forms of the Fourier transform \hat{f} and the inverse Fourier transform \check{f} of a Lebesgue integrable function $f \in L^1(\mathbb{R}^d)$

$$\hat{f}(\xi) := \left(\frac{1}{\sqrt{2\pi}} \right)^d \int_{\mathbb{R}^d} f(x) e^{-i(x, \xi)} dx, \quad \check{f}(\xi) := \left(\frac{1}{\sqrt{2\pi}} \right)^d \int_{\mathbb{R}^d} f(x) e^{i(x, \xi)} dx, \quad \xi \in \mathbb{R}^d.$$

Here (x, ξ) is the standard inner product on \mathbb{R}^d . Thus, one sees that

$$\text{sinc}(x - y) = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} e^{-i(\xi, x-y)} d\xi, \quad x, y \in \mathbb{R}^d. \quad (2.4)$$

Thus $\mathcal{W} := L^2([-\pi, \pi]^d)$ and $\Phi(x) := (\frac{1}{\sqrt{2\pi}})^d e^{-i(\xi, x)}$, $x \in \mathbb{R}^d$ satisfy (2.2). Lemma 2.1 tells that \mathcal{H}_K is the space of continuous square integrable functions on \mathbb{R}^d whose Fourier transforms are supported on $[-\pi, \pi]^d$ and the inner product on \mathcal{H}_K inherits from that of $L^2(\mathbb{R}^d)$. This is well-known. We use it to illustrate the application of Lemma 2.1.

Given two kernels K, G on a prescribed input space X , the corresponding RKHS $\mathcal{H}_K, \mathcal{H}_G$ can usually be identified by Lemma 2.1. The theme of the paper is the set inclusion relation $\mathcal{H}_K \subseteq \mathcal{H}_G$. As point evaluations are continuous on RKHS, it was observed in [1] that if $\mathcal{H}_K \subseteq \mathcal{H}_G$ then the identity operator from \mathcal{H}_K into \mathcal{H}_G is bounded. We shall denote by $\beta(K, G)$ the operator norm of this embedding. A characterization of $\mathcal{H}_K \subseteq \mathcal{H}_G$ was also established in [1]. Following [1], we write $K \ll G$ if $G - K$ remains a kernel on X .

Lemma 2.2 [1] *Let K, G be two kernels on X , $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if there exists a nonnegative constant $\lambda \geq 0$ such that $K \ll \lambda G$.*

Provided that $\mathcal{H}_K \subseteq \mathcal{H}_G$, we shall denote by $\lambda(K, G)$ the infimum of the set of positive constants λ such that $K \ll \lambda G$. If $\mathcal{H}_K \not\subseteq \mathcal{H}_G$ then we make the convention that $\lambda(K, G) = +\infty$. We first make a simple observation about the two quantities $\beta(K, G)$ and $\lambda(K, G)$.

Proposition 2.3 *Let K, G be two kernels on X with $\mathcal{H}_K \subseteq \mathcal{H}_G$ then $\beta(K, G) = \sqrt{\lambda(K, G)}$ and $K \ll \lambda(K, G)G$.*

Proof: It was proved in [1] that for two kernels K, L on X , $K \ll L$ if and only if $\mathcal{H}_K \subseteq \mathcal{H}_L$ and $\|f\|_{\mathcal{H}_L} \leq \|f\|_{\mathcal{H}_K}$ for all $f \in \mathcal{H}_K$. Note by Lemma 2.1 that \mathcal{H}_G and $\mathcal{H}_{\lambda G}$ share common elements and for all $f \in \mathcal{H}_G$ that

$$\|f\|_{\mathcal{H}_G} = \sqrt{\lambda} \|f\|_{\mathcal{H}_{\lambda G}}.$$

Combing these two facts, we get that for all $\lambda > 0$ that $K \ll \lambda G$ if and only if $\mathcal{H}_K \subseteq \mathcal{H}_G$ and $\|f\|_{\mathcal{H}_G} \leq \sqrt{\lambda} \|f\|_{\mathcal{H}_K}$ for all $f \in \mathcal{H}_K$. Thus, if $K \ll \lambda G$ then $\beta(K, G) \leq \sqrt{\lambda}$. It follows that $\beta(K, G) \leq \sqrt{\lambda(K, G)}$. On the other hand, if $\beta > \beta(K, G)$ then there exists some $f \in \mathcal{H}_K$ for which either $f \notin \mathcal{H}_G$ or $\|f\|_{\mathcal{H}_G} > \beta \|f\|_{\mathcal{H}_K}$. It implies that $K \ll \beta^2 G$ does not hold. As a consequence, $\lambda(K, G) \leq \beta^2$. We hence have that $\sqrt{\lambda(K, G)} \leq \beta(K, G)$, leading to the equality

$$\beta(K, G) = \sqrt{\lambda(K, G)},$$

which in turn implies that $K \ll \lambda(K, G)G$. □

We next present another characterization of the inclusion relation in terms of feature maps of reproducing kernels.

Theorem 2.4 *Let K, G be two kernels on X with the feature map $\Phi_1 : X \rightarrow \mathcal{W}_1$ and $\Phi_2 : X \rightarrow \mathcal{W}_2$, respectively. If $\overline{\text{span}} \Phi_1(X) = \mathcal{W}_1$ and $\overline{\text{span}} \Phi_2(X) = \mathcal{W}_2$ then $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if there exists a bounded linear operator $T : \mathcal{W}_2 \rightarrow \mathcal{W}_1$ such that*

$$T\Phi_2(x) = \Phi_1(x), \quad x \in X. \tag{2.5}$$

Moreover, the inclusion is nontrivial if and only if the adjoint operator T^ of T is not surjective.*

Proof: The result can be proved by similar arguments as those in Theorems 6 and 7 of [23]. □

By the above theorem, the particular choices

$$\mathcal{W}_1 := \mathcal{H}_K, \quad \Phi_1(x) := K(x, \cdot), \quad \mathcal{W}_2 := \mathcal{H}_G, \quad \Phi_2(x) = G(x, \cdot), \quad x \in X$$

yields that $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if there exists a bounded operator $L : \mathcal{H}_G \rightarrow \mathcal{H}_K$ such that $LG(x, \cdot) = K(x, \cdot)$ for all $x \in X$. We remark that this result in the special case when X is a countable dense subset of \mathbb{R}^d was proved in [6].

3 Translation Invariant Kernels and Radial Basis Functions

Translation invariant kernels are the most widely-used class of reproducing kernels on the Euclidean space. A kernel K on \mathbb{R}^d is said to be *translation invariant* if

$$K(x - a, y - a) = K(x, y) \text{ for all } x, y, a \in \mathbb{R}^d.$$

There is a celebrated characterization of continuous translation invariant kernels on \mathbb{R}^d due to Bochner [3]. The result is usually referred to as the Bochner theorem. Denote by $\mathcal{B}(\mathbb{R}^d)$ the set of all the finite positive Borel measures on \mathbb{R}^d . The characterization states that continuous translation invariant kernels on \mathbb{R}^d are exactly the Fourier transform of finite positive Borel measures in $\mathcal{B}(\mathbb{R}^d)$. Thus we shall consider the inclusion relation $\mathcal{H}_K \subseteq \mathcal{H}_G$ for two translation invariant kernels K, G of the form

$$K(x, y) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} d\mu(\xi), \quad x, y \in \mathbb{R}^d, \quad (3.1)$$

and

$$G(x, y) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} d\nu(\xi), \quad x, y \in \mathbb{R}^d, \quad (3.2)$$

where $\mu, \nu \in \mathcal{B}(\mathbb{R}^d)$.

Let μ, ν be two finite Borel measures on a topological space Y . Recall that μ is said to be *absolutely continuous* with respect to ν , denoted as $\mu \ll \nu$, if μ vanishes on Borel subsets of Y with zero ν measure. When $\mu \ll \nu$, $d\mu/d\nu$ is a Borel measurable function on Y such that

$$\mu(A) = \int_A \frac{d\mu}{d\nu}(x) d\nu(x) \quad \text{for all Borel subsets } A \subseteq Y.$$

We denote by $L_\nu^\infty(Y)$ the space of Borel measurable functions on Y with the norm

$$\|f\|_{L_\nu^\infty(Y)} := \inf\{M > 0 : \nu(\{t \in Y : |f(t)| > M\}) = 0\} < +\infty.$$

For later use, we also denote by $L_\nu^2(Y)$ the Hilbert space of Borel measurable functions on Y such that

$$\|f\|_{L_\nu^2(Y)} := \left(\int_Y |f(t)|^2 d\nu(t) \right)^{1/2} < +\infty.$$

Proposition 3.1 *Let K, G be two continuous translation invariant kernels on \mathbb{R}^d given by (3.1) and (3.2). Then $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if $\mu \ll \nu$ and $d\mu/d\nu \in L_\nu^\infty(\mathbb{R}^d)$. In the case that $\mathcal{H}_K \subseteq \mathcal{H}_G$,*

$$\lambda(K, G) = \left\| \frac{d\mu}{d\nu} \right\|_{L_\nu^\infty(\mathbb{R}^d)}. \quad (3.3)$$

Proof: By Lemma 2.2, $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if there exists some $\lambda \geq 0$ such that $\lambda G - K$ is a kernel on \mathbb{R}^d . Note that for all $\lambda \geq 0$, $\lambda G - K$ is still translation invariant. Therefore, by the Bochner theorem, $K \leq \lambda G$ if and only if $\lambda\nu - \mu \in \mathcal{B}(\mathbb{R}^d)$, which happens if and only if $\mu \ll \nu$ and $d\mu/d\nu$ is bounded by λ almost everywhere on \mathbb{R}^d with respect to ν . We hence get that $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if $\mu \ll \nu$ and $d\mu/d\nu \in L_\nu^\infty(\mathbb{R}^d)$. When $\mu \ll \nu$ and $d\mu/d\nu \in L_\nu^\infty(\mathbb{R}^d)$, it is clear that (3.3) holds. \square

We pay special attention to the situation when the Borel measures in (3.1) and (3.2) are absolutely continuous with respect to the Lebesgue measure. In this case, by the Radon-Nikodym theorem, K, G are the Fourier transform of nonnegative Lebesgue integrable functions on \mathbb{R}^d .

Corollary 3.2 *Let u, v be nonnegative functions in $L^1(\mathbb{R}^d)$ and let K, G be defined by*

$$K(x, y) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} u(\xi) d\xi, \quad G(x, y) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} v(\xi) d\xi, \quad x, y \in \mathbb{R}^d. \quad (3.4)$$

Then $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if the set $\{t \in \mathbb{R}^d : u(t) > 0, v(t) = 0\}$ has Lebesgue measure zero and u/v is essentially bounded on $\{t \in \mathbb{R}^d : v(t) > 0\}$, in which case $\lambda(K, G)$ equals the essential upper bound of u/v on $\{t \in \mathbb{R}^d : v(t) > 0\}$. In particular, if v is positive almost everywhere on \mathbb{R}^d then $\mathcal{H}_K \subseteq \mathcal{H}_G$ if and only if $u/v \in L^\infty(\mathbb{R}^d)$, in which case $\lambda(K, G) = \|u/v\|_{L^\infty(\mathbb{R}^d)}$.

An important class of translation invariant kernels on \mathbb{R}^d are given by radial basis functions. Those are reproducing kernels of the form

$$K_d(x, y) = g(\|x - y\|), \quad x, y \in \mathbb{R}^d, \quad (3.5)$$

where g is a single-variate function on $\mathbb{R}_+ := [0, +\infty)$ and $\|\cdot\|$ is the standard Euclidean norm on \mathbb{R}^d . The following well-known characterizations of kernels of the form (3.5) are due to Schoenberg [15].

For each $d \in \mathbb{N}$, denote by $d\omega_d$ and ω_d the area element and total area of the unit sphere of \mathbb{R}^d , respectively. Also set

$$\Omega_d(|x|) := \frac{1}{\omega_d} \int_{\|\xi\|=1} e^{i(x, \xi)} d\omega_d(\xi), \quad x \in \mathbb{R}^d.$$

Lemma 3.3 *Let g be a function on \mathbb{R}_+ . Then (3.5) defines a reproducing kernel on \mathbb{R}^d if and only if there is a finite positive Borel measure μ on \mathbb{R}_+ such that*

$$K_d(x, y) = \int_0^\infty \Omega_d(t\|x - y\|) d\mu(t), \quad x, y \in \mathbb{R}^d. \quad (3.6)$$

Furthermore, equation (3.5) defines a reproducing kernel K_d on \mathbb{R}^d for all $d \in \mathbb{N}$ if and only if

$$K_d(x, y) = \int_0^\infty e^{-t\|x-y\|^2} d\mu(t), \quad x, y \in \mathbb{R}^d \quad (3.7)$$

for some finite positive Borel measure μ on \mathbb{R}_+ .

Notice that both $\text{span}\{\Omega_d(tr) : r > 0\}$ and $\text{span}\{e^{-tr} : r > 0\}$ are dense in $C_0(\mathbb{R}_+)$, the space of continuous functions on \mathbb{R}_+ vanishing at infinity equipped with the maximum norm. By this fact and Lemma 3.3, one may use arguments similar to those in the proof of Proposition 3.1 to get the following characterizations of the inclusion relation of RKHS of kernels of the form (3.5).

Proposition 3.4 *Let μ, ν be two finite positive Borel measures on \mathbb{R}_+ , let K_d be given by (3.6) and set*

$$G_d(x, y) := \int_0^\infty \Omega_d(t\|x - y\|) d\nu(t), \quad x, y \in \mathbb{R}^d. \quad (3.8)$$

Then $\mathcal{H}_{K_d} \subseteq \mathcal{H}_{G_d}$ if and only if $\mu \ll \nu$ and $d\mu/d\nu \in L^\infty_\nu(\mathbb{R}_+)$, in which case $\lambda(K_d, G_d) = \|d\mu/d\nu\|_{L^\infty_\nu(\mathbb{R}_+)}$. If K_d is given by (3.7) and G_d is defined by

$$G_d(x, y) = \int_0^\infty e^{-t\|x-y\|^2} d\nu(t), \quad x, y \in \mathbb{R}^d \quad (3.9)$$

then $\mathcal{H}_{K_d} \subseteq \mathcal{H}_{G_d}$ for all $d \in \mathbb{N}$ and $\{\lambda(K_d, G_d) : d \in \mathbb{N}\}$ is bounded if and only if $\mu \ll \nu$ and $d\mu/d\nu \in L^\infty_\nu(\mathbb{R}_+)$, in which case $\sup\{\lambda(K_d, G_d) : d \in \mathbb{N}\} = \|d\mu/d\nu\|_{L^\infty_\nu(\mathbb{R}_+)}$.

One may specify statements in the above proposition to the case when μ, ν are absolutely continuous with respect to the Lebesgue measure on \mathbb{R}_+ to get results similar to those in Corollary 3.2, which we shall not state here.

We next turn to the main purpose of this section, which is to explore the inclusion relations among the RKHS of six commonly used translation invariant kernels in machine learning and other areas of applied mathematics. To apply the characterizations established above, we present those kernels in the form they appear in the characterization of Bochner or Schoenberg:

- the Gaussian kernel

$$G_\gamma(x, y) = \exp\left(-\frac{\|x - y\|^2}{\gamma}\right) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} g_\gamma(\xi) d\xi, \quad x, y \in \mathbb{R}^d, \quad \gamma > 0. \quad (3.10)$$

where

$$g_\gamma(\xi) := \left(\frac{\sqrt{\gamma}}{2\sqrt{\pi}}\right)^d \exp\left(-\frac{\gamma\|\xi\|^2}{4}\right), \quad \xi \in \mathbb{R}^d.$$

- the ℓ^1 -norm exponential kernel

$$E_{\sigma_1}(x, y) = \exp\left(-\frac{\|x - y\|_1}{\sigma_1}\right) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} \varphi_{\sigma_1}(\xi) d\xi, \quad x, y \in \mathbb{R}^d, \quad \sigma_1 > 0, \quad (3.11)$$

where $\|x\|_1 := \sum_{j=1}^d |x_j|$, $x = (x_j : j \in \mathbb{N}_d) \in \mathbb{R}^d$ and

$$\varphi_{\sigma_1}(\xi) := \frac{\sigma_1^d}{\pi^d} \prod_{j=1}^d \frac{1}{1 + \sigma_1^2 \xi_j^2}, \quad \xi \in \mathbb{R}^d.$$

- the ℓ^2 -norm exponential kernel

$$\mathcal{E}_{\sigma_2}(x, y) = \exp(-\|x - y\|) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} \psi_{\sigma_2}(\xi) d\xi, \quad x, y \in \mathbb{R}^d, \quad \sigma_2 > 0, \quad (3.12)$$

where

$$\psi_{\sigma_2}(\xi) := \frac{\Gamma(\frac{d+1}{2})}{\pi^{\frac{d+1}{2}}} \frac{\sigma_2^d}{(1 + \sigma_2^2 \|\xi\|^2)^{\frac{d+1}{2}}}, \quad \xi \in \mathbb{R}^d. \quad (3.13)$$

Here, Γ denotes the Gamma function and the Fourier transform is identified by the Poisson kernel (see, for example, [20], page 61).

- the inverse multiquadrics

$$M_\beta(x, y) := \frac{1}{(1 + \|x - y\|^2)^\beta} = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} m_\beta(\xi) d\xi, \quad x, y \in \mathbb{R}^d, \quad \beta > 0, \quad (3.14)$$

where

$$m_\beta(\xi) := \frac{1}{(2\sqrt{\pi})^d} \frac{1}{\Gamma(\beta)} \int_0^\infty t^{\beta-\frac{d}{2}-1} \exp\left(-\frac{\|\xi\|^2}{4t} - t\right) dt, \quad \xi \in \mathbb{R}^d. \quad (3.15)$$

This formulation can be obtained by combining Theorem 7.15 in [22] and the Fourier transform of the Gaussian function.

- the B-spline kernel

$$B_p(x, y) := \prod_{j=1}^d B_p(x_j - y_j) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} b_p(\xi) d\xi, \quad x, y \in \mathbb{R}^d, \quad p \in 2\mathbb{N}, \quad (3.16)$$

where B_p denotes the p -th order cardinal B-spline, and with $\text{sinc } \frac{1}{2}(t) := \frac{\sin(\frac{t}{2})}{\frac{t}{2}}$, $t \in \mathbb{R}$,

$$b_p(\xi) := \frac{1}{(2\pi)^d} \prod_{j=1}^d (\text{sinc } \frac{1}{2}(\xi_j))^p, \quad \xi \in \mathbb{R}^d.$$

- the ANOVA kernel

$$A_\tau(x, y) := \sum_{j=1}^d \exp\left(-\frac{|x_j - y_j|^2}{\tau}\right) = \int_{\mathbb{R}^d} e^{i(x-y, \xi)} a_\tau(\xi) d\xi, \quad x, y \in \mathbb{R}^d, \quad \tau > 0, \quad (3.17)$$

where

$$a_\tau(\xi) := \frac{\sqrt{\tau}}{2\sqrt{\pi}} \left(\sum_{j=1}^d \exp\left(-\frac{\tau \xi_j^2}{4}\right) \right), \quad \xi \in \mathbb{R}^d.$$

Among those kernels, the Gaussian kernel, the ℓ^2 -norm exponential kernel, and the inverse multiquadrics are radial basis functions. We also give their representation by the Laplace transform below:

- the Gaussian kernel

$$G_\gamma(x, y) = \exp\left(-\frac{\|x - y\|^2}{\gamma}\right) = \int_0^\infty e^{-\|x-y\|^2 t} d\delta_{\gamma^{-1}}(t), \quad (3.18)$$

where δ_t denotes the unit measure supported at the singleton $\{t\}$.

- the ℓ^2 -norm exponential kernel

$$\mathcal{E}_{\sigma_2}(x, y) = \exp\left(-\frac{\|x - y\|}{\sigma_2}\right) = \frac{1}{2\sigma_2\sqrt{\pi}} \int_0^\infty e^{-\|x-y\|^2 t} \exp\left(-\frac{1}{4\sigma_2^2 t}\right) \frac{1}{t^{3/2}} dt, \quad x, y \in \mathbb{R}^d. \quad (3.19)$$

This equation is derived from the identity (see [20], page 61) that

$$e^{-r} = \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-r^2/4s} \frac{e^{-s}}{\sqrt{s}} ds, \quad r > 0.$$

- the inverse multiquadrics (see [22], page 95)

$$M_\beta(x, y) = \frac{1}{(1 + \|x - y\|^2)^\beta} = \frac{1}{\Gamma(\beta)} \int_0^\infty e^{-\|x-y\|^2 t} t^{\beta-1} e^{-t} dt, \quad x, y \in \mathbb{R}^d. \quad (3.20)$$

As a straightforward application of Corollary 3.2, we have the following inclusion relations between the RKHS of kernels of the same kind.

Proposition 3.5 *The following statements hold true:*

(1) For $0 < \gamma_1 < \gamma_2$, $\mathcal{H}_{G_{\gamma_2}} \subseteq \mathcal{H}_{G_{\gamma_1}}$ with

$$\lambda(G_{\gamma_2}, G_{\gamma_1}) = \left(\frac{\gamma_2}{\gamma_1} \right)^{\frac{d}{2}},$$

but $\mathcal{H}_{G_{\gamma_1}} \not\subseteq \mathcal{H}_{G_{\gamma_2}}$.

(2) For $0 < \sigma_1 < \sigma_2$, $\mathcal{H}_{E_{\sigma_1}} = \mathcal{H}_{E_{\sigma_2}}$ with

$$\lambda(E_{\sigma_1}, E_{\sigma_2}) = \lambda(E_{\sigma_2}, E_{\sigma_1}) = \left(\frac{\sigma_2}{\sigma_1} \right)^d.$$

(3) For $0 < \sigma_1 < \sigma_2$, $\mathcal{H}_{\mathcal{E}_{\sigma_1}} = \mathcal{H}_{\mathcal{E}_{\sigma_2}}$ with

$$\lambda(\mathcal{E}_{\sigma_1}, \mathcal{E}_{\sigma_2}) = \frac{\sigma_2}{\sigma_1}, \quad \lambda(\mathcal{E}_{\sigma_2}, \mathcal{E}_{\sigma_1}) = \left(\frac{\sigma_2}{\sigma_1} \right)^d.$$

(4) For $p, q \in 2\mathbb{N}$ with $p < q$, $\mathcal{H}_{B_q} \subseteq \mathcal{H}_{B_p}$ with $\lambda(B_q, B_p) = 1$, but $\mathcal{H}_{B_p} \not\subseteq \mathcal{H}_{B_q}$.

(5) For $0 < \tau_1 < \tau_2$, $\mathcal{H}_{A_{\tau_2}} \subseteq \mathcal{H}_{A_{\tau_1}}$ with $\lambda(A_{\tau_2}, A_{\tau_1}) = \sqrt{\frac{\tau_2}{\tau_1}}$, but $\mathcal{H}_{A_{\tau_1}} \not\subseteq \mathcal{H}_{A_{\tau_2}}$.

The inclusion relation for the RKHS of two inverse multiquadrics is more involved and is separated below.

Theorem 3.6 *Let β_1, β be two distinct positive constants. There holds $\mathcal{H}_{M_{\beta_1}} \subseteq \mathcal{H}_{M_{\beta_2}}$ if and only if $\frac{d}{2} < \beta_1 < \beta_2$.*

Proof: Suppose first that $\beta_1 > \beta_2$. By the same technique used in Theorem 6.13, [22] and equation (3.15), one obtains for all $\beta > 0$ that

$$m_\beta(\xi) = \frac{2^{1-\beta}}{(\sqrt{2\pi})^d \Gamma(\beta)} \|\xi\|^{\beta-\frac{d}{2}} K_{\beta-\frac{d}{2}}(\|\xi\|), \quad \xi \neq 0, \quad (3.21)$$

where K_ν , $\nu \in \mathbb{R}$ is the modified Bessel functions defined by

$$K_\nu(r) := \int_0^\infty e^{-r \cosh t} \cosh(\nu t) dt, \quad r > 0.$$

We use the estimates (see, [22], pages 52-53) about K_ν that there exists a constant C_ν depending on ν only such that

$$K_\nu(r) \geq C_\nu \frac{e^{-r}}{\sqrt{r}}, \quad r \geq 1 \quad (3.22)$$

and that

$$K_\nu(r) \leq \sqrt{2\pi} \frac{e^{-r}}{\sqrt{r}} \exp\left(\frac{|\nu|^2}{2r}\right), \quad r > 0. \quad (3.23)$$

Combining equations (3.21), (3.22), and (3.23), we obtain for $\beta_1 > \beta_2$ that

$$\frac{m_{\beta_1}(\xi)}{m_{\beta_2}(\xi)} \geq \frac{C_{\beta_1 - \frac{d}{2}} 2^{\beta_2 - \beta_1} \Gamma(\beta_2)}{\sqrt{2\pi} \Gamma(\beta_1)} \|\xi\|^{\beta_1 - \beta_2} \exp\left(-\frac{|\beta_2 - \frac{d}{2}|^2}{2\|\xi\|}\right), \quad \|\xi\| \geq 1. \quad (3.24)$$

Since the right hand side above goes to infinity as $\|\xi\| \rightarrow \infty$, we get by Corollary 3.2 that $\mathcal{H}_{M_{\beta_1}} \not\subseteq \mathcal{H}_{M_{\beta_2}}$ when $\beta_1 > \beta_2$.

By monotone convergence theorem, we have by equation (3.15) for all $\beta > 0$ that

$$\lim_{\xi \rightarrow 0} m_{\beta}(\xi) = \begin{cases} +\infty, & \text{if } \beta \leq \frac{d}{2}, \\ \frac{1}{(2\sqrt{\pi})^d} \frac{\Gamma(\beta - \frac{d}{2})}{\Gamma(\beta)} < +\infty, & \text{if } \beta > \frac{d}{2}. \end{cases} \quad (3.25)$$

Therefore, if $\beta_1 \leq \frac{d}{2} < \beta_2$ then $m_{\beta_1}(\xi)/m_{\beta_2}(\xi)$ is unbounded on a neighborhood of the origin. As a consequence, $\mathcal{H}_{M_{\beta_1}} \not\subseteq \mathcal{H}_{M_{\beta_2}}$ in this case.

Suppose that $\frac{d}{2} < \beta_1 < \beta_2$. Then by (3.25), $m_{\beta_1}(\xi)/m_{\beta_2}(\xi)$ is bounded on a neighborhood of the origin. Also, by (3.24),

$$\lim_{\|\xi\| \rightarrow \infty} \frac{m_{\beta_1}(\xi)}{m_{\beta_2}(\xi)} = 0.$$

As $m_{\beta_1}(\xi)/m_{\beta_2}(\xi)$ is continuous on $\mathbb{R}^d \setminus \{0\}$, it is hence bounded therein. By Corollary 3.2, $\mathcal{H}_{M_{\beta_1}} \subseteq \mathcal{H}_{M_{\beta_2}}$ when $\frac{d}{2} < \beta_1 < \beta_2$.

We now discuss the last case that $\beta_1 < \beta_2 \leq \frac{d}{2}$. We shall show that in this case $\mathcal{H}_{M_{\beta_1}} \not\subseteq \mathcal{H}_{M_{\beta_2}}$ by proving that $m_{\beta_1}(\xi)/m_{\beta_2}(\xi)$ is unbounded on a neighborhood of the origin. To this end, let $\|\xi\| \leq 1$ and use the change of variables $t = \|\xi\|^2 s$ in (3.15) to get that

$$\frac{m_{\beta_1}(\xi)}{m_{\beta_2}(\xi)} = \|\xi\|^{2(\beta_1 - \beta_2)} \frac{\Gamma(\beta_2)}{\Gamma(\beta_1)} \frac{\int_0^\infty s^{\beta_1 - \frac{d}{2} - 1} \exp\left(-\frac{1}{4s} - \|\xi\|^2 s\right) ds}{\int_0^\infty s^{\beta_2 - \frac{d}{2} - 1} \exp\left(-\frac{1}{4s} - \|\xi\|^2 s\right) ds}. \quad (3.26)$$

Thus, if $\beta_2 < \frac{d}{2}$ then we have for $\|\xi\| \leq 1$ that

$$\frac{m_{\beta_1}(\xi)}{m_{\beta_2}(\xi)} \geq \|\xi\|^{2(\beta_1 - \beta_2)} \frac{\Gamma(\beta_2)}{\Gamma(\beta_1)} \frac{\int_0^\infty s^{\beta_1 - \frac{d}{2} - 1} \exp\left(-\frac{1}{4s} - s\right) ds}{\int_0^\infty s^{\beta_2 - \frac{d}{2} - 1} \exp\left(-\frac{1}{4s}\right) ds}.$$

The right hand side above is unbounded when $\|\xi\| \rightarrow 0$. When $\beta_2 = \frac{d}{2}$, we estimate that

$$\int_0^\infty s^{\beta_2 - \frac{d}{2} - 1} \exp\left(-\frac{1}{4s} - \|\xi\|^2 s\right) ds \leq \int_0^1 \frac{1}{s} e^{-\frac{1}{4s}} ds + \int_1^\infty \frac{1}{s} e^{-\|\xi\|^2 s} ds.$$

A change of variables $\|\xi\|^2 s = t$ then yields for $\|\xi\| < 1$ that

$$\int_1^\infty \frac{1}{s} e^{-\|\xi\|^2 s} ds = \int_{\|\xi\|^2}^\infty \frac{1}{t} e^{-t} dt \leq \int_{\|\xi\|^2}^1 \frac{1}{t} dt + \int_1^\infty \frac{e^{-t}}{t} dt = -2 \ln(\|\xi\|) + \int_1^\infty \frac{e^{-t}}{t} dt.$$

Combining the above two equations with (3.26) yields that

$$\frac{m_{\beta_1}(\xi)}{m_{\beta_2}(\xi)} \geq \|\xi\|^{2(\beta_1-\beta_2)} \frac{\Gamma(\beta_2)}{\Gamma(\beta_1)} \frac{\int_0^\infty s^{\beta_1-\frac{d}{2}-1} \exp\left(-\frac{1}{4s}-s\right) ds}{\int_0^1 \frac{1}{s} e^{-\frac{1}{4s}} ds + \int_1^\infty \frac{e^{-t}}{t} dt - 2\ln(\|\xi\|)}, \quad \|\xi\| < 1.$$

The right hand side above goes to infinity as $\|\xi\| \rightarrow 0$. The proof is complete. \square

The main purpose of this section is to explore the inclusion relationships among RKHS of different kinds of translation invariant kernels given above. We present the results in the form of a table.

Theorem 3.7 *Let $p \in 2\mathbb{N}$ and $\gamma, \sigma_1, \sigma_2, \beta, \tau$ be positive constants. The following inclusion relations of RKHS hold true.*

	\mathcal{H}_{B_p}	\mathcal{H}_{G_γ}	$\mathcal{H}_{E_{\sigma_1}}$	$\mathcal{H}_{\mathcal{E}_{\sigma_2}}$	\mathcal{H}_{M_β}	\mathcal{H}_{A_τ}
\mathcal{H}_{B_p}	$=$	$\not\subseteq$	\subseteq	\subseteq iff $p \geq d+1$	$\not\subseteq$	$\not\subseteq$
\mathcal{H}_{G_γ}	$\not\subseteq$	$=$	\subseteq	\subseteq	\subseteq	\subseteq iff $\gamma \geq \tau$
$\mathcal{H}_{E_{\sigma_1}}$	$\not\subseteq$	$\not\subseteq$	$=$	$\not\subseteq$ if $d \geq 2$	$\not\subseteq$	$\not\subseteq$
$\mathcal{H}_{\mathcal{E}_{\sigma_2}}$	$\not\subseteq$	$\not\subseteq$	$\not\subseteq$ if $d \geq 2$	$=$	$\not\subseteq$	$\not\subseteq$
\mathcal{H}_{M_β}	$\not\subseteq$	$\not\subseteq$	\subseteq iff $\beta > \frac{d}{2}$	\subseteq iff $\beta > \frac{d}{2}$	$=$	$\not\subseteq$
\mathcal{H}_{A_τ}	$\not\subseteq$	$\not\subseteq$ if $d \geq 2$	$\not\subseteq$	$\not\subseteq$	$\not\subseteq$	$=$

We break the task of proving this result into several steps as follows.

- (i) For any dimension $d \in \mathbb{N}$ and parameters $p \in 2\mathbb{N}$, $\gamma, \tau > 0$, $\mathcal{H}_{B_p} \not\subseteq \mathcal{H}_K$ and $\mathcal{H}_K \not\subseteq \mathcal{H}_{B_p}$ for $K = G_\gamma$ or $K = A_\tau$.

Proof: We first discuss the case when $K = G_\gamma$. It is clear that b_p/g_γ is unbounded on \mathbb{R}^d . By Corollary 3.2, $\mathcal{H}_{B_p} \not\subseteq \mathcal{H}_{G_\gamma}$. On the other hand, b_p possesses zeros on \mathbb{R}^d while g_γ is everywhere positive. As they are both continuous, there does not exist a positive constant $\lambda > 0$ such that $g_\gamma(\xi) \leq \lambda^2 b_p(\xi)$ for almost every $\xi \in \mathbb{R}^d$. As a consequence, we obtain by Corollary 3.2 that $\mathcal{H}_{G_\gamma} \not\subseteq \mathcal{H}_{B_p}$. The other case when $K = A_\tau$ can be handled in a similar way. \square

- (ii) For any $d \in \mathbb{N}$, $\sigma_2 > 0$ and $p \in 2\mathbb{N}$, $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{B_p}$. There holds $\mathcal{H}_{B_p} \subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$ if and only if $p \geq d+1$, in which case

$$\lambda(B_p, \mathcal{E}_{\sigma_2}) \leq \frac{2^{p-d}}{\sigma_2^d \pi^{\frac{d-1}{2}}} \frac{(1 + \sigma_2^2 d)^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}. \quad (3.27)$$

Proof: The function ψ_{σ_2} in (3.12) is continuous and positive everywhere on \mathbb{R}^d . By arguments used before, $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{B_p}$. Assume that $p < d+1$. We choose $\xi_1 = (2n+1)\pi$ and $\xi_j = 0$ for $j \geq 2$ to get that $b_p(\xi) = O(n^{-p})$ while $\psi_{\sigma_2}(\xi) = O(n^{-(d+1)})$ as n tends to infinity. Therefore, b_p/ψ_{σ_2} is unbounded on \mathbb{R}^d , implying that $\mathcal{H}_{B_p} \not\subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$.

Suppose that $p \geq d+1$. If $\|\xi\|_\infty := \max\{|\xi_j| : j \in \mathbb{N}_d\} \leq 1$ then

$$b_p(\xi) \leq \frac{1}{(2\pi)^d}, \quad \psi_{\sigma_2}(\xi) \geq \frac{\Gamma(\frac{d+1}{2})}{\pi^{\frac{d+1}{2}}} \frac{\sigma_2^d}{(1 + \sigma_2^2 d)^{\frac{d+1}{2}}}.$$

It follows that

$$\frac{b_p(\xi)}{\psi_{\sigma_2}(\xi)} \leq \frac{1}{(2\sigma_2\pi)^d} \frac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})} (1 + \sigma_2^2 d)^{\frac{d+1}{2}}, \quad \|\xi\|_\infty \leq 1. \quad (3.28)$$

When $\|\xi\|_\infty \geq 1$,

$$b_p(\xi) \leq \frac{1}{(2\pi)^d} \frac{2^p}{\|\xi\|_\infty^p},$$

which implies by $p \geq d+1$ that for $\|\xi\|_\infty \geq 1$,

$$\begin{aligned} \frac{b_p(\xi)}{\psi_{\sigma_2}(\xi)} &\leq \frac{2^p}{(2\pi)^d} \frac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})} \frac{(1 + \sigma_2^2 d \|\xi\|_\infty^2)^{\frac{d+1}{2}}}{\sigma_2^d \|\xi\|_\infty^p} \leq \frac{2^p}{(2\sigma_2\pi)^d} \frac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})} \left(\frac{1}{\|\xi\|_\infty^2} + \sigma_2^2 d \right)^{\frac{d+1}{2}} \\ &\leq \frac{2^p}{(2\sigma_2\pi)^d} \frac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})} (1 + \sigma_2^2 d)^{\frac{d+1}{2}}. \end{aligned}$$

By Corollary 3.2, the above inequality together with (3.28) proves (3.27). \square

(iii) For any $d \in \mathbb{N}$, $\sigma_1 > 0$ and $p \in 2\mathbb{N}$, $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{B_p}$. There holds $\mathcal{H}_{B_p} \subseteq \mathcal{H}_{E_{\sigma_1}}$ and

$$\lambda(B_p, E_{\sigma_1}) \leq 2^d \left(\sigma_1 + \frac{1}{\sigma_1} \right)^d. \quad (3.29)$$

Proof: The relation $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{B_p}$ follows from that φ_{σ_1} is positive and continuous everywhere on \mathbb{R}^d . Using an estimate method similar to that in (ii), we get that

$$(\text{sinc } \tfrac{1}{2}(t))^p (1 + \sigma_1^2 t^2) \leq (\text{sinc } \tfrac{1}{2}(t))^2 (1 + \sigma_1^2 t^2) \leq 4(1 + \sigma_1^2) \text{ for all } t \in \mathbb{R},$$

which combined with the explicit form of b_p and φ_{σ_1} leads to (3.29). \square

(iv) For any $d \in \mathbb{N}$, $\sigma_1 > 0$ and $\gamma > 0$, $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{G_\gamma}$. There holds $\mathcal{H}_{G_\gamma} \subseteq \mathcal{H}_{E_{\sigma_1}}$ and

$$\lambda(G_\gamma, E_{\sigma_1}) \leq \left(\max(1, \frac{4\sigma_1^2}{\gamma}) \frac{\sqrt{\gamma\pi}}{2\sigma_1} \right)^d. \quad (3.30)$$

Proof: It is clear that $\varphi_{\sigma_1}/g_\gamma$ is unbounded on \mathbb{R}^d . By Corollary 3.2, $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{G_\gamma}$. On the other hand, one has that

$$\frac{g_\gamma(\xi)}{\varphi_{\sigma_1}(\xi)} = \left(\frac{\sqrt{\gamma\pi}}{2\sigma_1} \right)^d \exp\left(-\frac{\gamma\|\xi\|^2}{4}\right) \prod_{j=1}^d (1 + \sigma_1^2 \xi_j^2), \quad \xi \in \mathbb{R}^d,$$

which together with the observation that

$$(1 + \sigma_1^2 \xi_j^2) \leq \max(1, \frac{4\sigma_1^2}{\gamma}) \exp\left(\frac{\gamma \xi_j^2}{4}\right), \quad \xi_j \in \mathbb{R}$$

proves (3.30). \square

(v) For any $d \in \mathbb{N}$, $\sigma_1 > 0$ and $\gamma > 0$, $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{G_\gamma}$. There holds $\mathcal{H}_{G_\gamma} \subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$ and

$$\lambda(G_\gamma, \mathcal{E}_{\sigma_2}) \leq \left(\max(1, \frac{(2d+2)\sigma_2^2}{\gamma}) \right)^{\frac{d+1}{2}} \left(\frac{\sqrt{\gamma}}{2\sigma_2} \right)^d \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2})}. \quad (3.31)$$

However, $\lambda(G_\gamma, \mathcal{E}_{\sigma_2})$ does not have a common upper bound as d varies on \mathbb{N} .

Proof: As ψ_{σ_2}/g_γ is clearly unbounded on \mathbb{R}^d , $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{G_\gamma}$. We then estimate that for all $\xi \in \mathbb{R}^d$,

$$(1+\sigma_2^2\|\xi\|^2)^{\frac{d+1}{2}} \leq \left(\max(1, \frac{(2d+2)\sigma_2^2}{\gamma}) \right)^{\frac{d+1}{2}} (1+\frac{\gamma\|\xi\|^2}{2d+2})^{\frac{d+1}{2}} \leq \left(\max(1, \frac{(2d+2)\sigma_2^2}{\gamma}) \right)^{\frac{d+1}{2}} \exp(\frac{\gamma\|\xi\|^2}{4}),$$

which immediately implies that $g_\gamma(\xi)/\psi_{\sigma_2}(\xi)$ is bounded by the right hand side of (3.31). Equation (3.31) now follows from Corollary 3.2.

To prove the third claim, we use the Laplace transform representations (3.18) and (3.19). One observes that the Gaussian kernel G_γ corresponds to the delta measure $\delta_{\gamma^{-1}}$, which is singular with respect to the Lebesgue measure while \mathcal{E}_{σ_2} is represented by the Borel measure

$$\frac{1}{2\sigma_2\sqrt{\pi}} \exp(-\frac{1}{4\sigma_2^2 t}) \frac{1}{t^{3/2}} dt,$$

which is absolutely continuous with respect to the Lebesgue measure. Thus, $\delta_{\gamma^{-1}}$ is not absolutely continuous with respect to the above measure. By Proposition 3.4, $\lambda(G_\gamma, \mathcal{E}_{\sigma_2})$ does not have a common upper bound as the dimension d varies on \mathbb{N} . \square

(vi) For any $d \geq 2$, $\sigma_1 > 0$ and $\sigma_2 > 0$, $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{E_{\sigma_1}}$ and $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$.

Proof: We first let $\xi_1 = n$ and $\xi_j = 0$ for $j \geq 2$ to get that $\varphi_{\sigma_1}(\xi) = O(n^{-2})$ and $\psi_{\sigma_2}(\xi) = O(n^{-(d+1)})$ as n tends to infinity. As $d \geq 2$, $\varphi_{\sigma_1}(\xi)/\psi_{\sigma_2}(\xi)$ is unbounded on \mathbb{R}^d , implying that $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$. The choice $\xi_j = n$ for all $j \in \mathbb{N}_d$ yields that $\varphi_{\sigma_1}(\xi) = O(n^{-2d})$ and $\psi_{\sigma_2}(\xi) = O(n^{-(d+1)})$ as $n \rightarrow \infty$. Therefore, $\psi_{\sigma_2}(\xi)/\varphi_{\sigma_1}(\xi)$ is unbounded on \mathbb{R}^d . It implies that $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{E_{\sigma_1}}$. \square

(vii) For any $d \geq 2$, $\sigma_1, \sigma_2, \tau > 0$, $\mathcal{H}_{A_\tau} \not\subseteq \mathcal{H}_K$ and $\mathcal{H}_K \not\subseteq \mathcal{H}_{A_\tau}$ for either $K = E_{\sigma_1}$ or $K = \mathcal{E}_{\sigma_2}$.

Proof: We discuss $K = E_{\sigma_1}$ only as the other case can be dealt with similarly. Choosing $\xi_j = n$ for all $j \in \mathbb{N}_d$ yields that $\varphi_{\sigma_1}(\xi)/a_\tau(\xi) \rightarrow \infty$ as $n \rightarrow \infty$. The other choice $\xi_1 = n$ and $\xi_j = 0$ for $j \geq 2$ tells that $a_\tau(\xi)/\varphi_{\sigma_1}(\xi) \rightarrow \infty$ as $n \rightarrow \infty$. Therefore, neither $\varphi_{\sigma_1}/a_\tau$ nor $a_\tau/\varphi_{\sigma_1}$ is bounded on \mathbb{R}^d . The result now follows from Corollary 3.2. \square

(viii) For any $d \geq 2$, $\gamma, \tau > 0$, $\mathcal{H}_{A_\tau} \not\subseteq \mathcal{H}_{G_\gamma}$. There holds $\mathcal{H}_{G_\gamma} \subseteq \mathcal{H}_{A_\tau}$ if and only if $\gamma \geq \tau$, in which case

$$\lambda(G_\gamma, A_\tau) = \frac{\sqrt{\gamma}^d}{d\sqrt{\tau}(2\sqrt{\pi})^{d-1}}. \quad (3.32)$$

Proof: That $\mathcal{H}_{A_\tau} \not\subseteq \mathcal{H}_{G_\gamma}$ can be proved in a way similar to that in (vii). If $\gamma < \tau$ then we set $\xi_j = n$ for all $j \in \mathbb{N}_d$ to see that $g_\gamma(\xi)/a_\tau(\xi) \rightarrow \infty$ as $n \rightarrow \infty$. Thus, $\mathcal{H}_{G_\gamma} \not\subseteq \mathcal{H}_{A_\tau}$ in this case.

Suppose that $\gamma \geq \tau$. We get for all $\xi \in \mathbb{R}^d$ that

$$\frac{g_\gamma(\xi)}{a_\tau(\xi)} = \frac{\sqrt{\gamma}^d}{\sqrt{\tau}(2\sqrt{\pi})^{d-1}} \frac{\exp(-\frac{\gamma\|\xi\|^2}{4})}{\sum_{j=1}^d \exp(-\frac{\tau|\xi_j|^2}{4})},$$

which together with the observation that

$$\exp(-\frac{\gamma\|\xi\|^2}{4}) \leq \exp(-\frac{\tau|\xi_j|^2}{4}) \text{ for all } j \in \mathbb{N}_d$$

implies that

$$\frac{g_\gamma(\xi)}{a_\tau(\xi)} \leq \frac{\sqrt{\gamma}^d}{d\sqrt{\tau}(2\sqrt{\pi})^{d-1}} \text{ for all } \xi \in \mathbb{R}^d.$$

As the equality is achieved at $\xi = 0$, we obtain (3.32). \square

(ix) For any $d \in \mathbb{N}$, $\sigma_2, \beta > 0$, $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{M_\beta}$. There holds $\mathcal{H}_{M_\beta} \subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$ if and only if $\beta > \frac{d}{2}$.

Proof: By (3.13) and (3.15), we have for all $\xi \in \mathbb{R}^d$ that

$$\frac{m_\beta(\xi)}{\psi_{\sigma_2}(\xi)} = \frac{1}{(2\sigma_2\sqrt{\pi})^d} \frac{\pi^{\frac{d+1}{2}}}{\Gamma(\beta)\Gamma(\frac{d+1}{2})} \int_0^\infty t^{\beta-\frac{d}{2}-1} (1 + \sigma_2^2\|\xi\|^2)^{\frac{d+1}{2}} \exp\left(-\frac{\|\xi\|^2}{4t} - t\right) dt. \quad (3.33)$$

Note that when $\|\xi\| \geq 1$,

$$\exp\left(-\frac{\|\xi\|^2}{4t} - t\right) = \exp\left(-\frac{\|\xi\|^2}{8t} - \frac{t}{2}\right) \exp\left(-\frac{\|\xi\|^2}{8t} - \frac{t}{2}\right) \leq \exp\left(-\frac{\|\xi\|}{2}\right) \exp\left(-\frac{1}{8t} - \frac{t}{2}\right).$$

Thus, for $\|\xi\| \geq 1$

$$\int_0^\infty t^{\beta-\frac{d}{2}-1} (1 + \sigma_2^2\|\xi\|^2)^{\frac{d+1}{2}} \exp\left(-\frac{\|\xi\|^2}{4t} - t\right) dt \leq \frac{(1 + \sigma_2^2\|\xi\|^2)^{\frac{d+1}{2}}}{\exp(\frac{\|\xi\|}{2})} \int_0^\infty t^{\beta-\frac{d}{2}-1} \exp\left(-\frac{1}{8t} - \frac{t}{2}\right) dt.$$

We hence get that $m_\beta(\xi)/\psi_{\sigma_2}(\xi) \rightarrow 0$ as $\|\xi\| \rightarrow \infty$. It implies that $\mathcal{H}_{\mathcal{E}_{\sigma_2}} \not\subseteq \mathcal{H}_{M_\beta}$.

To prove the rest of the claims, one first sees by the Lebesgue dominated convergence theorem that m_β/ψ_{σ_2} is continuous on $\mathbb{R}^d \setminus \{0\}$. We also have that $m_\beta(\xi)/\psi_{\sigma_2}(\xi) \rightarrow 0$ as $\|\xi\| \rightarrow \infty$. For these two reasons, m_β/ψ_{σ_2} is essentially bounded on \mathbb{R}^d if and only if it is bounded on a neighborhood of the origin. If $\beta > \frac{d}{2}$, we observe that when $\|\xi\| \leq 1$,

$$\int_0^\infty t^{\beta-\frac{d}{2}-1} (1 + \sigma_2^2\|\xi\|^2)^{\frac{d+1}{2}} \exp\left(-\frac{\|\xi\|^2}{4t} - t\right) dt \leq (1 + \sigma_2^2)^{\frac{d+1}{2}} \int_0^\infty t^{\beta-\frac{d}{2}-1} e^{-t} dt < +\infty,$$

which implies that m_β/ψ_{σ_2} is essentially bounded on \mathbb{R}^d when $\beta > \frac{d}{2}$. We hence get by Corollary 3.2 that $\mathcal{H}_{M_\beta} \subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$ in this case. When $\beta \leq \frac{d}{2}$, by the monotone convergence theorem,

$$\lim_{\|\xi\| \rightarrow 0} \int_0^\infty t^{\beta-\frac{d}{2}-1} (1 + \sigma_2^2\|\xi\|^2)^{\frac{d+1}{2}} \exp\left(-\frac{\|\xi\|^2}{4t} - t\right) dt = \int_0^\infty t^{\beta-\frac{d}{2}-1} e^{-t} dt = +\infty.$$

It follows from the above equation that $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{\mathcal{E}_{\sigma_2}}$ when $\beta \leq \frac{d}{2}$. \square

- (x) For any $d \in \mathbb{N}$, $\sigma_1, \beta > 0$, $\mathcal{H}_{E_{\sigma_1}} \not\subseteq \mathcal{H}_{M_\beta}$. There holds $\mathcal{H}_{M_\beta} \subseteq \mathcal{H}_{E_{\sigma_1}}$ if and only if $\beta > \frac{d}{2}$.

Proof: The proof is similar to that for (ix). \square

- (xi) For any $d \in \mathbb{N}$, $p \in 2\mathbb{N}$, $\beta > 0$, $\mathcal{H}_{B_p} \not\subseteq \mathcal{H}_{M_\beta}$ and $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{B_p}$.

Proof: As m_β is positive and continuous on $\mathbb{R}^d \setminus \{0\}$ while b_p has zeros on $\mathbb{R}^d \setminus \{0\}$, $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{B_p}$. That $\mathcal{H}_{B_p} \not\subseteq \mathcal{H}_{M_\beta}$ can be proved by arguments similar to those in (ix). \square

- (xii) For any $d \in \mathbb{N}$, $\gamma, \beta > 0$, $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{G_\gamma}$ but $\mathcal{H}_{G_\gamma} \subseteq \mathcal{H}_{M_\beta}$. The quantity $\lambda(G_\gamma, M_\beta)$ does not have a common upper bound as d varies on \mathbb{N} .

Proof: We start with the observation that

$$\begin{aligned} \frac{m_\beta(\xi)}{g_\gamma(\xi)} &= \frac{1}{\Gamma(\beta)\gamma^{\frac{d}{2}}} \int_0^\infty t^{\beta-\frac{d}{2}-1} \exp\left(\frac{\gamma\|\xi\|^2}{4}\right) \exp\left(-\frac{\|\xi\|^2}{4t} - t\right) dt \\ &\geq \frac{1}{\Gamma(\beta)\gamma^{\frac{d}{2}}} \int_{\frac{2}{\gamma}}^\infty t^{\beta-\frac{d}{2}-1} \exp\left(\frac{\gamma\|\xi\|^2}{4} - \frac{\|\xi\|^2}{4t}\right) e^{-t} dt \\ &\geq \frac{1}{\Gamma(\beta)\gamma^{\frac{d}{2}}} \exp\left(\frac{\gamma\|\xi\|^2}{8}\right) \int_{\frac{2}{\gamma}}^\infty t^{\beta-\frac{d}{2}-1} e^{-t} dt. \end{aligned}$$

Therefore, $m_\beta(\xi)/g_\gamma(\xi)$ tends to infinity as $\|\xi\| \rightarrow \infty$. Consequently, $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{G_\gamma}$.

We also notice by the monotone convergence theorem that

$$\lim_{\|\xi\| \rightarrow 0} \frac{m_\beta(\xi)}{g_\gamma(\xi)} = \frac{1}{\Gamma(\beta)\gamma^{\frac{d}{2}}} \int_0^\infty t^{\beta-\frac{d}{2}-1} e^{-t} dt > 0.$$

As $\frac{m_\beta(\xi)}{g_\gamma(\xi)}$ is continuous and positive everywhere on $\mathbb{R}^d \setminus \{0\}$, the above two estimates imply that there exists some positive constant λ such that

$$\frac{m_\beta(\xi)}{g_\gamma(\xi)} \geq \lambda \text{ for all } \xi \in \mathbb{R}^d \setminus \{0\}.$$

We hence conclude that $\mathcal{H}_{G_\gamma} \subseteq \mathcal{H}_{M_\beta}$. Recall (3.18) and (3.20). Since G_γ and M_β are respectively represented by measures singular and absolutely continuous with respect to the Lebesgue measure, $\lambda(G_\gamma, M_\beta)$ does not have a common upper bound for $d \in \mathbb{N}$. \square

- (xiii) For any $d \in \mathbb{N}$, $\tau, \beta > 0$, $\mathcal{H}_{A_\tau} \not\subseteq \mathcal{H}_{M_\beta}$ and $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{A_\tau}$.

Proof: Firstly, we see for the choice $\xi_1 = n$, $\xi_j = 0$, $j \geq 2$ that

$$\lim_{n \rightarrow \infty} m_\beta(\xi) = 0 \text{ while } \lim_{n \rightarrow \infty} a_\tau(\xi) = \frac{(d-1)\sqrt{\tau}}{2\sqrt{\pi}}.$$

As a result, $\mathcal{H}_{A_\tau} \not\subseteq \mathcal{H}_{M_\beta}$. Secondly, arguments similar to those in (xii) shows that for the choice $\xi_j = n$, $j \in \mathbb{N}_d$

$$\lim_{n \rightarrow \infty} \frac{m_\beta(\xi)}{a_\tau(\xi)} = +\infty,$$

which implies that $\mathcal{H}_{M_\beta} \not\subseteq \mathcal{H}_{A_\tau}$. \square

We close this section with the sinc kernel (2.4).

Corollary 3.8 *There holds for all $\gamma > 0$ and $d \in \mathbb{N}$ that*

$$\lambda(\text{sinc}, G_\gamma) = \frac{\exp(\frac{d\gamma\pi^2}{4})}{(\gamma\pi)^{\frac{d}{2}}}, \quad \lambda(\text{sinc}, A_\tau) = \frac{\sqrt{\pi} \exp(\frac{\tau\pi^2}{4})}{2^{d-1}\pi^d d \sqrt{\tau}}. \quad (3.34)$$

Consequently, $\mathcal{H}_{\text{sinc}} \subseteq \mathcal{H}_K$ for $K = E_{\sigma_1}, \mathcal{E}_{\sigma_2}$, and M_β .

Proof: Equation (3.34) follows from a straightforward calculation. \square

4 Hilbert-Schmidt Kernels

By Mercer's theorem [12], Hilbert-Schmidt kernels represent a large class of reproducing kernels. They were recently used to construct multiscale kernels based on wavelets [14]. We introduce the general form of Hilbert-Schmidt kernels.

Let a be a nonnegative function on \mathbb{N} and set $a_n := a(n)$, $n \in \mathbb{N}$. We denote by $\ell_a^2(\mathbb{N})$ the Hilbert space of functions c on \mathbb{N} such that

$$\|c\|_{\ell_a^2(\mathbb{N})} := \left(\sum_{n=1}^{\infty} a_n |c_n|^2 \right)^{1/2} < +\infty.$$

Its inner product is given by

$$(c, d)_{\ell_a^2(\mathbb{N})} := \sum_{n=1}^{\infty} a_n c_n \overline{d_n}, \quad c, d \in \ell_a^2(\mathbb{N}).$$

Suppose that we have a sequence of functions ϕ_n , $n \in \mathbb{N}$, on the input space X , such that for each $x \in X$ the function $\Phi(x)$ defined on \mathbb{N} as

$$\Phi(x)(n) := \phi_n(x), \quad n \in \mathbb{N} \quad (4.1)$$

belongs to $\ell_a^2(\mathbb{N})$. The Hilbert-Schmidt kernel K_a associate with a is given as

$$K_a(x, y) := (\Phi(x), \Phi(y))_{\ell_a^2(\mathbb{N})} = \sum_{n=1}^{\infty} a_n \phi_n(x) \overline{\phi_n(y)}, \quad x, y \in X. \quad (4.2)$$

Now suppose that there exists another nonnegative function b on \mathbb{N} such that $\Phi(x) \in \ell_b^2(\mathbb{N})$ for all $x \in X$. Set

$$K_b(x, y) := (\Phi(x), \Phi(y))_{\ell_b^2(\mathbb{N})} = \sum_{n=1}^{\infty} b_n \phi_n(x) \overline{\phi_n(y)}, \quad x, y \in X. \quad (4.3)$$

We shall characterize $\mathcal{H}_{K_a} \subseteq \mathcal{H}_{K_b}$ in terms of a and b .

Proposition 4.1 *Suppose that b is nontrivial, and $\text{span}\{\Phi(x) : x \in X\}$ is dense in both $\ell_a^2(\mathbb{N})$ and $\ell_b^2(\mathbb{N})$. Then $\mathcal{H}_{K_a} \subseteq \mathcal{H}_{K_b}$ if and only if there is a constant $\lambda > 0$ such that $a_n \leq \lambda b_n$ for all $n \in \mathbb{N}$. In this case,*

$$\lambda(K_a, K_b) = \sup \left\{ \frac{a_n}{b_n} : n \in \mathbb{N}, b_n > 0 \right\}. \quad (4.4)$$

Proof: By Lemma 2.1, the space \mathcal{H}_{K_a} consists of functions of the form

$$f_c(x) := (c, \Phi(x))_{\ell_a^2(\mathbb{N})} = \sum_{n=1}^{\infty} c_n a_n \overline{\phi_n(x)}, \quad x \in X, \quad c \in \ell_a^2(\mathbb{N}) \quad (4.5)$$

with the norm

$$\|f_c\|_{\mathcal{H}_{K_a}} = \|c\|_{\ell_a^2(\mathbb{N})}.$$

Similarly, one has the structure of the space \mathcal{H}_{K_b} .

Suppose that there exists some constant $\lambda > 0$ such that $a_n \leq \lambda b_n$ for all $n \in \mathbb{N}$. Let c be an arbitrary but fixed element in $\ell_a^2(\mathbb{N})$ and set

$$\tilde{c}_n := \begin{cases} 0, & \text{if } a_n = 0, \\ \frac{a_n c_n}{b_n}, & \text{otherwise.} \end{cases}$$

One sees that $\tilde{c} \in \ell_b^2(\mathbb{N})$ and that $(\tilde{c}, \Phi(\cdot))_{\ell_b^2(\mathbb{N})} = f_c$. Thus, $f_c \in \mathcal{H}_{K_b}$, implying that $\mathcal{H}_{K_a} \subseteq \mathcal{H}_{K_b}$. Another observation is that

$$\|f_c\|_{\mathcal{H}_{K_b}}^2 = \sum_{n \in \mathbb{N}, a_n \neq 0} \left| \frac{a_n c_n}{b_n} \right|^2 b_n \leq \sup\{a_n/b_n : n \in \mathbb{N}, a_n \neq 0\} \|f_c\|_{\mathcal{H}_{K_a}}^2.$$

Moreover, for any $k \in \mathbb{N}$ with $a_k > 0$, the particular choice $c(n) := \delta_{n,k}$, $n \in \mathbb{N}$, where $\delta_{n,k}$ denotes the Kronecker delta, yields that

$$\|f_c\|_{\mathcal{H}_{K_b}}^2 = \frac{a_k^2}{b_k} = \frac{a_k}{b_k} \|f_c\|_{\mathcal{H}_{K_a}}^2.$$

The above two equations together imply by Proposition 2.3 that

$$\lambda(K_a, K_b) = \beta(K_a, K_b)^2 = \sup \left\{ \frac{a_n}{b_n} : n \in \mathbb{N}, b_n > 0 \right\}.$$

Conversely, suppose that $\mathcal{H}_{K_a} \subseteq \mathcal{H}_{K_b}$. As the embedding operator is bounded, there exists $\lambda > 0$ such that $\|f\|_{\mathcal{H}_{K_b}} \leq \lambda \|f\|_{\mathcal{H}_{K_a}}$ for all $f \in \mathcal{H}_{K_a}$. For any $k \in \mathbb{N}$ with $a_k > 0$, we still choose $c(n) := \delta_{n,k}$, $n \in \mathbb{N}$ to get from $f_c \in \mathcal{H}_{K_b}$ that $b_k > 0$ and that

$$\|f_c\|_{\mathcal{H}_{K_b}}^2 = \frac{a_k^2}{b_k^2} b_k \leq \lambda \|f_c\|_{\mathcal{H}_{K_a}}^2 = \lambda a_k,$$

which implies that $a_k \leq \lambda b_k$. The proof is complete. \square

Before we give examples of inclusion relations for Hilbert-Schmidt kernels by Proposition 4.1, we remark that Proposition 4.1 actually leads to a characterization of Hilbert-Schmidt kernels.

Theorem 4.2 *Let r be a function on \mathbb{N} . Suppose that $\Phi(x) \in \ell_{|r|}^2(\mathbb{N})$ for all $x \in X$ and $\text{span } \Phi(X)$ is dense in $\ell_{|r|}^2(\mathbb{N})$. Then*

$$K_r(x, y) := \sum_{n=1}^{\infty} r_n \phi_n(x) \overline{\phi_n(y)}, \quad x, y \in X \quad (4.6)$$

defines a kernel on X if and only if $r_n \geq 0$ for each $n \in \mathbb{N}$.

Proof: The sufficiency is well-known. We prove the necessity by contradiction. Assume that K_r given by (4.6) is a kernel but $r_{j_0} < 0$ for some $j_0 \in \mathbb{N}$. Then we introduce two nonnegative functions a and b on \mathbb{N} by setting

$$a_n = \begin{cases} 2|r_n|, & n \neq j_0, \\ -r_{j_0}, & n = j_0. \end{cases}$$

and

$$b_n = \begin{cases} 2|r_n| + r_n, & n \neq j_0, \\ 0, & n = j_0. \end{cases}$$

Then it is clear that $\Phi(x) \in \ell_a^2(\mathbb{N})$ and $\Phi(x) \in \ell_b^2(\mathbb{N})$ for all $x \in X$. Moreover, $\text{span } \Phi(X)$ is dense in $\ell_a^2(\mathbb{N})$ and $\ell_b^2(\mathbb{N})$ as it is in $\ell_{|r|}^2(\mathbb{N})$. Therefore, K_a and K_b are Hilbert-Schmidt kernels on X . Note that $K_b - K_a = K_r$. By the assumption, $K_a \ll K_b$. Thus by Proposition 4.1, there exists some $\lambda > 0$ such that $a_n \leq \lambda b_n$ for all $n \in \mathbb{N}$. Especially when $n = j_0$, we have $-r_{j_0} \leq \lambda 0 = 0$, contradicting that $r_{j_0} < 0$. \square

As an application of the above theorem, we discuss an important and celebrated result which was proved before by rather sophisticated mathematical analysis [16]. Suppose that the power series

$$\sum_{n=0}^{\infty} a_n z^n, \quad z \in \mathbb{C}$$

has a positive convergence radius r . Then by Corollary 4.2 or [16],

$$K(x, y) := \sum_{n=0}^{\infty} a_n (x, y)^n, \quad x, y \in \mathbb{R}^d, \quad \|x\|, \|y\| < r^{1/2}$$

is a reproducing kernel on $\{x \in \mathbb{R}^d : \|x\| < r^{1/2}\}$ if and only if $a_n \geq 0$ for all $n \geq 0$.

We close this section with a few examples that fall into the consideration of Proposition 4.1. We shall not state the results explicitly as they would just be repetition of those in Proposition 4.1.

- (Discrete Exponential Kernels) Let $t_n, n \in \mathbb{N}$ be a sequence of pairwise distinct points in \mathbb{R}^d and let a, b be two nonnegative functions in $\ell^1(\mathbb{N})$. The associated discrete exponential kernels are given by

$$K_a(x, y) := \sum_{n=1}^{\infty} a_n e^{i(x-y, t_n)}, \quad K_b(x, y) := \sum_{n=1}^{\infty} b_n e^{i(x-y, t_n)}, \quad x, y \in \mathbb{R}^d.$$

Useful examples of discrete exponential kernels including the periodic kernels (see, for example, [17], page 103). We present three instances below. Let γ, σ be positive constants and $\alpha > d$. Define

$$\tilde{G}_\gamma(x, y) := \sum_{n \in \mathbb{Z}^d} e^{i(x-y, n)} e^{-\gamma \|n\|^2}, \quad x, y \in [0, 2\pi]^d,$$

$$\tilde{E}_\sigma(x, y) := \sum_{n \in \mathbb{Z}^d} e^{i(x-y, n)} e^{-\sigma \|n\|}, \quad x, y \in [0, 2\pi]^d,$$

and

$$\tilde{P}_q(x, y) := \sum_{n \in \mathbb{Z}^d} e^{i(x-y, n)} \frac{1}{(1 + \|n\|)^\alpha}, \quad x, y \in [0, 2\pi]^d.$$

Then by Proposition 4.1, we clearly have that $\mathcal{H}_{\tilde{G}_\gamma} \subseteq \mathcal{H}_{\tilde{E}_\sigma} \subseteq \mathcal{H}_{\tilde{P}_q}$.

- (Polynomial Kernels) Let a, b be two nonnegative functions on $\mathbb{N}_+ := \mathbb{N} \cup \{0\}$. Suppose that $\sum_{n=0}^{\infty} a_n z^n$ and $\sum_{n=0}^{\infty} b_n z^n$ both have a positive convergence radius r_a and r_b , respectively. Then the polynomial kernels

$$K_a(x, y) := \sum_{n=0}^{\infty} a_n (x, y)^n,$$

$$K_b(x, y) := \sum_{n=0}^{\infty} b_n (x, y)^n,$$

on the input space $\{x \in \mathbb{R}^d : \|x\| < \min(\sqrt{r_a}, \sqrt{r_b})\}$ satisfy the assumptions of Proposition 4.1.

Especially, we have the following simple observation about finite polynomial kernels.

Proposition 4.3 (*Finite Polynomial Kernels*) Let $p, q \in \mathbb{N}$ and put

$$K_p(x, y) := (1 + (x, y))^p, \quad x, y \in \mathbb{R}^d \quad (4.7)$$

and

$$K_q(x, y) := (1 + (x, y))^q, \quad x, y \in \mathbb{R}^d. \quad (4.8)$$

Then $\mathcal{H}_{K_p} \subseteq \mathcal{H}_{K_q}$ if and only if $p \leq q$. When $p \leq q$, $\lambda(K_p, K_q) = 1$.

5 Constructional Results

In this section, we discuss the preservation of the inclusion relation of RKHS under various operations with the corresponding kernels. We start with some trivial observations from Lemma 2.2.

Proposition 5.1 Let K_1, K_2, G_1, G_2, K, G be reproducing kernels on the input space X . Then the following results hold true:

i.) If $\mathcal{H}_{K_1} \subseteq \mathcal{H}_{G_1}$ and $\mathcal{H}_{K_2} \subseteq \mathcal{H}_{G_2}$ then $\mathcal{H}_{K_1+K_2} \subseteq \mathcal{H}_{G_1+G_2}$ and

$$\lambda(K_1 + K_2, G_1 + G_2) \leq \max(\lambda(K_1, G_1), \lambda(K_2, G_2)).$$

ii.) Especially, if \mathcal{H}_{K_1} and \mathcal{H}_{K_2} are both contained in \mathcal{H}_G then $\mathcal{H}_{K_1+K_2} \subseteq \mathcal{H}_G$ and

$$\lambda(K_1 + K_2, G) \leq \lambda(K_1, G) + \lambda(K_2, G).$$

iii.) If $\mathcal{H}_K \subseteq \mathcal{H}_G$ then for all $a, b > 0$, $\mathcal{H}_{aK} \subseteq \mathcal{H}_{bG}$ and

$$\lambda(aK, bG) = \frac{a}{b} \lambda(K, G).$$

We next turn to the product of two kernels by first examining the more general tensor product of kernels. Let K, G be two kernels on X . The tensor product $K \otimes G$ of K, G is a new kernel on the extended input space $X \times X$ defined by

$$(K \otimes G)(\mathbf{x}, \mathbf{y}) := K(\mathbf{x}_1, \mathbf{y}_1)G(\mathbf{x}_2, \mathbf{y}_2), \quad \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2), \quad \mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in X \times X.$$

For further discussion, we shall make use of the Schur product theorem [11]. For two square matrices A, B of the same size, we denote by $A \circ B$ the Hadamard product of A, B , that is, $A \circ B$ is formed by pairwise multiplying elements from A and B . The Schur product theorem asserts that the Hadamard product of two positive semi-definite matrices is still positive semi-definite.

Proposition 5.2 *Let K_1, K_2, G_1, G_2 be kernels on X . If $\mathcal{H}_{K_1} \subseteq \mathcal{H}_{G_1}$ and $\mathcal{H}_{K_2} \subseteq \mathcal{H}_{G_2}$ then $\mathcal{H}_{K_1 \otimes K_2} \subseteq \mathcal{H}_{G_1 \otimes G_2}$ and*

$$\lambda(K_1 \otimes K_2, G_1 \otimes G_2) \leq \lambda(K_1, G_1)\lambda(K_2, G_2).$$

Proof: For notational simplicity, put $\lambda_1 := \lambda(K_1, G_1)$ and $\lambda_2 := \lambda(K_2, G_2)$. We shall show that $K_1 \otimes K_2 \ll \lambda_1 \lambda_2 G_1 \otimes G_2$ by definition. Let $\mathbf{z} := \{\mathbf{x}^j : j \in \mathbb{N}_n\}$ be a finite set of pairwise distinct points in $X \times X$. Set $\mathbf{z}_1 := \{\mathbf{x}_1^j : j \in \mathbb{N}_n\}$ and $\mathbf{z}_2 := \{\mathbf{x}_2^j : j \in \mathbb{N}_n\}$. We observe that

$$(G_1 \otimes G_2)[\mathbf{z}] = G_1[\mathbf{z}_1] \circ G_2[\mathbf{z}_2], \quad (K_1 \otimes K_2)[\mathbf{z}] = K_1[\mathbf{z}_1] \circ K_2[\mathbf{z}_2].$$

By Proposition 2.3, $K_1 \ll \lambda_1 G_1$ and $K_2 \ll \lambda_2 G_2$. As a result, $\lambda_1 G_1[\mathbf{z}_1] - K_1[\mathbf{z}_1]$ and $\lambda_2 G_2[\mathbf{z}_2] - K_2[\mathbf{z}_2]$ are both positive semi-definite. We now compute that

$$\begin{aligned} \lambda_1 \lambda_2 (G_1 \otimes G_2)[\mathbf{z}] - (K_1 \otimes K_2)[\mathbf{z}] &= \lambda_1 \lambda_2 G_1[\mathbf{z}_1] \circ G_2[\mathbf{z}_2] - K_1[\mathbf{z}_1] \circ K_2[\mathbf{z}_2] \\ &= (K_1[\mathbf{z}_1] + (\lambda_1 G_1[\mathbf{z}_1] - K_1[\mathbf{z}_1])) \circ (K_2[\mathbf{z}_2] + (\lambda_2 G_2[\mathbf{z}_2] - K_2[\mathbf{z}_2])) - K_1[\mathbf{z}_1] \circ K_2[\mathbf{z}_2] \\ &= K_1[\mathbf{z}_1] \circ (\lambda_2 G_2[\mathbf{z}_2] - K_2[\mathbf{z}_2]) + (\lambda_1 G_1[\mathbf{z}_1] - K_1[\mathbf{z}_1]) \circ K_2[\mathbf{z}_2] \\ &\quad + (\lambda_1 G_1[\mathbf{z}_1] - K_1[\mathbf{z}_1]) \circ (\lambda_2 G_2[\mathbf{z}_2] - K_2[\mathbf{z}_2]). \end{aligned}$$

By the Schur product theorem, the three matrices in the last step above are all positive semi-definite. Therefore, $K_1 \otimes K_2 \ll \lambda_1 \lambda_2 G_1 \otimes G_2$. The proof is complete. \square

Corollary 5.3 *Let K_1, K_2, G_1, G_2 be kernels on X . If $\mathcal{H}_{K_1} \subseteq \mathcal{H}_{G_1}$ and $\mathcal{H}_{K_2} \subseteq \mathcal{H}_{G_2}$ then $\mathcal{H}_{K_1 K_2} \subseteq \mathcal{H}_{G_1 G_2}$ and $\lambda(K_1 K_2, G_1 G_2) \leq \lambda(K_1, G_1)\lambda(K_2, G_2)$.*

Proof: The result follows from Proposition 5.2 and the observation that $K_1 K_2$ and $G_1 G_2$ can be viewed as the restriction of $K_1 \otimes K_2$ and $G_1 \otimes G_2$ on the diagonal of $X \times X$, respectively. \square

We next discuss limits of reproducing kernels. It is obvious by definition that the limit of a sequence of kernels remains a kernel [1].

Proposition 5.4 *Let $\{K_j : j \in \mathbb{N}\}$ and $\{G_j : j \in \mathbb{N}\}$ be two sequences of kernels on X that converge pointwise to kernels K and G , respectively. If $\mathcal{H}_{K_j} \subseteq \mathcal{H}_{G_j}$ for all $j \in \mathbb{N}$ and*

$$\sup\{\lambda(K_j, G_j) : j \in \mathbb{N}\} < +\infty \tag{5.1}$$

then $\mathcal{H}_K \subseteq \mathcal{H}_G$ and $\lambda(K, G) \leq \sup\{\lambda(K_j, G_j) : j \in \mathbb{N}\}$.

Proof: Suppose that $\mathcal{H}_{K_j} \subseteq \mathcal{H}_{G_j}$ for all $j \in \mathbb{N}$ and $\lambda := \sup\{\lambda(K_j, G_j) : j \in \mathbb{N}\} < +\infty$. Let \mathbf{x} be a finite set of sampling points in X and $y \in \mathbb{C}^n$ be fixed. Then as $K_j \ll \lambda G_j$, we have for all $j \in \mathbb{N}$ that

$$y^*(\lambda G_j[\mathbf{x}] - K_j[\mathbf{x}])y \geq 0.$$

Taking the limit as $j \rightarrow \infty$, we get that

$$y^*(\lambda G[\mathbf{x}] - K[\mathbf{x}])y \geq 0.$$

The proof is hence complete. \square

We remark that condition (5.1) may not be removed in the last proposition. For a simple contradictory example, we let G be an arbitrary nontrivial kernel on X and set $K_j := \frac{1}{j}G$ and $G_j := G$

for all $j \in \mathbb{N}$. It is clear that $\mathcal{H}_{K_j} = \mathcal{H}_{G_j} = \mathcal{H}_G$ for each $j \in \mathbb{N}$. But the limit of K_j is the trivial kernel. The inclusion relation is hence not kept in the limit kernels. The reason is that $\lambda(K_j, G_j) = j$ is unbounded.

With the help of Propositions 5.1, 5.4 and Corollary 5.3, we are ready to give a main result of this section. We shall use a fact proved in [8] that if K is a kernel and ϕ is analytic with nonnegative Taylor coefficients at the origin then $\phi(K)$ remains a kernel.

Theorem 5.5 *Let K and G be two kernels on X with $\mathcal{H}_K \subseteq \mathcal{H}_G$. Then $\mathcal{H}_{e^K} \subseteq \mathcal{H}_{e^{\lambda(K,G)G}}$. In particular, if $\lambda(K, G) \leq 1$ then $\mathcal{H}_{e^K} \subseteq \mathcal{H}_{e^G}$.*

Proof: We may assume that $\lambda(K, G) \leq 1$. Let $K_n := \sum_{j=0}^n \frac{K^j}{j!}$ and $G_n := \sum_{j=0}^n \frac{G^j}{j!}$ for each $n \in \mathbb{N}$. Then K_n, G_n converge pointwise to e^K and e^G , respectively. It also follows from Proposition 5.1 and Corollary 5.3 that

$$K_n \ll \left(\max_{0 \leq j \leq n} \lambda(K, G)^j \right) G_n.$$

It is clear that $\max_{0 \leq j \leq n} \lambda(K, G)^j$, $n \in \mathbb{N}$ are bounded by 1. The result now follows immediately from Proposition 5.4. \square

The arguments used in the above proof in fact are able to prove a more general result, which we present below.

Proposition 5.6 *Let K and G be two kernels on X with $\mathcal{H}_K \subseteq \mathcal{H}_G$. Suppose that ϕ is an analytic function with nonnegative Taylor coefficients a_j , $j \geq 0$ at the origin. Then $\mathcal{H}_{\phi(K)} \subseteq \mathcal{H}_{\phi(\lambda(K,G)G)}$. If, in addition, $\lambda(K, G) \leq 1$, then $\mathcal{H}_{\phi(K)} \subseteq \mathcal{H}_{\phi(G)}$.*

6 Equivalent Norm Inclusion

In this section, we investigate a special inclusion relation where an equivalence on the norms on the smaller space is imposed. Specifically, for two kernels K, G on X , we denote by $\mathcal{H}_K \lesssim \mathcal{H}_G$ if $\mathcal{H}_K \subseteq \mathcal{H}_G$ and there exists positive constants α, β such that

$$\alpha \|f\|_{\mathcal{H}_K} \leq \|f\|_{\mathcal{H}_G} \leq \beta \|f\|_{\mathcal{H}_K} \text{ for all } f \in \mathcal{H}_K. \quad (6.1)$$

For an existing kernel K , we call a kernel G a *weak refinement* of K if $\mathcal{H}_K \lesssim \mathcal{H}_G$. This is a relaxation of the refinement kernel defined in [24] and is expected to accommodate more examples of reproducing kernels.

We start our investigation with a characterization of the equivalent norm inclusion relation. The following result from [1] is needed.

Lemma 6.1 *Let K and G be kernels on X . Then there holds for all $f \in \mathcal{H}_{K+G}$ that*

$$\|f\|_{\mathcal{H}_{K+G}}^2 = \min\{\|f_1\|_{\mathcal{H}_K}^2 + \|f_2\|_{\mathcal{H}_G}^2 : f = f_1 + f_2, f_1 \in \mathcal{H}_K, f_2 \in \mathcal{H}_G\}.$$

Theorem 6.2 *Let K and G be kernels on X with $\mathcal{H}_K \subseteq \mathcal{H}_G$. Then $\mathcal{H}_K \lesssim \mathcal{H}_G$ if and only if there exists some constant $\delta > 0$ such that*

$$\|e\|_{\mathcal{H}_{\lambda(K,G)G-K}} \geq \delta \|e\|_{\mathcal{H}_K} \text{ for each } e \in \mathcal{H}_K \cap \mathcal{H}_{\lambda(K,G)G-K}. \quad (6.2)$$

Proof: For notational simplicity, put $L := \lambda(K, G)G - K$. By Proposition 2.3, L is a kernel on X . Suppose that condition (6.2) is satisfied. Note that for each $f \in \mathcal{H}_K \subseteq \mathcal{H}_G$ with the decomposition $f = f_1 + f_2$ where $f_1 \in \mathcal{H}_K$, $f_2 \in \mathcal{H}_L$, we have $f_2 \in \mathcal{H}_K \cap \mathcal{H}_L$. This together with $\mathcal{H}_K \subseteq \mathcal{H}_G = \mathcal{H}_{\lambda(K, G)G}$ implies by Lemma 6.1 that for all $f \in \mathcal{H}_K$

$$\begin{aligned} \|f\|_{\mathcal{H}_{\lambda(K, G)G}}^2 &= \min_{f=f_1+f_2} \{\|f_1\|_{\mathcal{H}_K}^2 + \|f_2\|_{\mathcal{H}_L}^2 : f_1 \in \mathcal{H}_K, f_2 \in \mathcal{H}_L\} \\ &\geq \min_{f=f_1+f_2} \{\|f_1\|_{\mathcal{H}_K}^2 + \delta^2 \|f_2\|_{\mathcal{H}_K}^2 : f_1 \in \mathcal{H}_K, f_2 \in \mathcal{H}_L\} \\ &\geq \min_{f=f_1+f_2} \min\{1, \delta^2\} \{\|f_1\|_{\mathcal{H}_K}^2 + \|f_2\|_{\mathcal{H}_K}^2 : f_1 \in \mathcal{H}_K, f_2 \in \mathcal{H}_L\} \\ &\geq \frac{1}{2} \min\{1, \delta^2\} \|f\|_{\mathcal{H}_K}^2. \end{aligned}$$

Recall that for all $f \in \mathcal{H}_G$,

$$\|f\|_{\mathcal{H}_G} = \sqrt{\lambda(K, G)} \|f\|_{\mathcal{H}_{\lambda(K, G)G}}.$$

By the above two equations and Proposition 2.3, we have for all $f \in \mathcal{H}_K$ that

$$\frac{1}{\sqrt{2}} \min\{1, \delta\} \sqrt{\lambda(K, G)} \|f\|_{\mathcal{H}_K} \leq \|f\|_{\mathcal{H}_G} \leq \sqrt{\lambda(K, G)} \|f\|_{\mathcal{H}_K}$$

in other words, $\mathcal{H}_K \lesssim \mathcal{H}_G$.

Conversely, suppose that $\mathcal{H}_K \lesssim \mathcal{H}_G$ but (6.2) does not hold for any $\delta > 0$. Then for each $n \in \mathbb{N}$, there exists $g_n \in \mathcal{H}_K \cap \mathcal{H}_L$ such that

$$\|g_n\|_{\mathcal{H}_L} \leq \frac{1}{n} \|g_n\|_{\mathcal{H}_K}. \quad (6.3)$$

Since $L \ll \lambda(K, G)G$, it follows from Lemmas 2.2 and 6.1 that $\mathcal{H}_L \subseteq \mathcal{H}_{\lambda(K, G)G}$ and

$$\|g_n\|_{\mathcal{H}_G} = \sqrt{\lambda(K, G)} \|g_n\|_{\mathcal{H}_{\lambda(K, G)G}} \leq \sqrt{\lambda(K, G)} \|g_n\|_{\mathcal{H}_L} \text{ for all } n \in \mathbb{N}. \quad (6.4)$$

Equations (6.3) and (6.4) imply that

$$\|g_n\|_{\mathcal{H}_G} \leq \frac{\sqrt{\lambda(K, G)}}{n} \|g_n\|_{\mathcal{H}_K} \text{ for all } n \in \mathbb{N},$$

contradicting (6.1). The proof is complete. \square

As an application of Theorem 6.2, we have the following example.

Proposition 6.3 *Consider the two finite polynomial kernels K_p, K_q defined by (4.7) and (4.8). Then $\mathcal{H}_{K_p} \lesssim \mathcal{H}_{K_q}$ if and only if $p \leq q$.*

Proof: By Proposition 4.3, $\mathcal{H}_{K_p} \subseteq \mathcal{H}_{K_q}$ if and only if $p \leq q$. Thus, if $\mathcal{H}_{K_p} \lesssim \mathcal{H}_{K_q}$ then $p \leq q$. Suppose that $p \leq q$. We introduce another kernel K on \mathbb{R}^d by setting

$$K(x, y) := \sum_{j=0}^p \binom{q}{j} (x, y)^j, \quad x, y \in \mathbb{R}^d.$$

Then by Proposition 4.1, $\mathcal{H}_K \subseteq \mathcal{H}_{K_q}$ and $\mathcal{H}_K = \mathcal{H}_{K_p}$. It is clear that $\mathcal{H}_K \cap \mathcal{H}_{K_q-K} = \{0\}$. By Theorem 6.2, $\mathcal{H}_K \lesssim \mathcal{H}_{K_q}$. As $\mathcal{H}_K = \mathcal{H}_{K_p}$, we have $\mathcal{H}_{K_p} \lesssim \mathcal{H}_{K_q}$. The proof is complete. \square

Before moving on, we make a simple observation that if two kernels K, G on X satisfy $\mathcal{H}_K \lesssim \mathcal{H}_G$ and $\mathcal{H}_K \neq \mathcal{H}_G$ then \mathcal{H}_K can not be dense in \mathcal{H}_G . For instances, given two Gaussian kernels $G_{\gamma_1}, G_{\gamma_2}$ with $\gamma_1 < \gamma_2$. As $\mathcal{H}_{G_{\gamma_2}} \subseteq \mathcal{H}_{G_{\gamma_1}}$ and $\mathcal{H}_{G_{\gamma_2}}$ is dense in but not equal to $\mathcal{H}_{G_{\gamma_1}}$, G_{γ_1} is not a weak refinement of G_{γ_2} .

The main purpose of this section is to present two characterizations of the equivalent norm inclusion that are widely applicable to translation invariant kernels and Hilbert-Schmidt kernels. As the study would be similar to that in [24], we shall omit the proof and examples.

Let μ, ν be two finite positive Borel measures on a topological space Y . Set

$$\omega := \frac{\mu + \nu}{2} + \frac{|\mu - \nu|}{2},$$

where $|\mu - \nu|$ denotes the total variation measure of $\mu - \nu$. Then μ and ν are absolutely continuous with respect to ω . Given a function $\phi : X \times Y \rightarrow \mathbb{C}$ such that $\phi(x, \cdot) \in L^2_\omega(Y)$ for all $x \in X$ and

$$\overline{\text{span}}\{\phi(x, \cdot) : x \in X\} = L^2_\omega(Y), \quad (6.5)$$

we introduce two kernels K_μ, K_ν on X by setting

$$K_\mu(x, y) := (\phi(x, \cdot), \phi(y, \cdot))_{L^2_\mu(Y)}, \quad K_\nu(x, y) := (\phi(x, \cdot), \phi(y, \cdot))_{L^2_\nu(Y)}, \quad x, y \in X. \quad (6.6)$$

Our task is to characterize the equivalent inclusion relation $\mathcal{H}_K \lesssim \mathcal{H}_G$ in terms of the measures μ and ν . To this end, we write $\mu \lesssim \nu$ if $\mu \ll \nu$ and there exist positive constants α, β such that $\alpha \leq d\mu/d\nu \leq \beta$ almost everywhere on $\{t \in Y : \frac{d\mu}{d\nu}(t) > 0\}$ with respect to ν .

The following characterization theorem can be proved by arguments similar to those in [24].

Theorem 6.4 *Suppose that $\phi : X \times Y \rightarrow \mathbb{C}$ satisfies (6.5) and K_μ, K_ν are defined by (6.6). Then $\mathcal{H}_\mu \lesssim \mathcal{H}_\nu$ if and only if $\mu \lesssim \nu$.*

The above theorem has a particular application to Hilbert-Schmidt kernels. For two nonnegative functions a, b on \mathbb{N} , we denote by $a \lesssim b$ if $\text{supp } a \subseteq \text{supp } b$ and there exist two positive constants α and β such that $\alpha a_n \leq b_n \leq \beta a_n$ for each $n \in \text{supp } a$. Here $\text{supp } a := \{n \in \mathbb{N} : a_n \neq 0\}$. Recall the definition of Hilbert-Schmidt kernels (4.2) and (4.3) through a sequence of functions (4.1).

Proposition 6.5 *Suppose that $\text{span}\{\Phi(x) : x \in X\}$ is dense in both $\ell^2_a(\mathbb{N})$ and $\ell^2_b(\mathbb{N})$. Then $\mathcal{H}_{K_a} \lesssim \mathcal{H}_{K_b}$ if and only if $a \lesssim b$.*

We want to reemphasize that our results, though similar to those in [24] for refinement of reproducing kernels, much increase the chance of refining an existing kernel. Taking polynomial kernels as an instance, for two such kernels

$$K(x, y) := \sum_{j=0}^N a_j(x, y)^j, \quad G(x, y) := \sum_{k=0}^M b_k(x, y)^k, \quad x, y \in \mathbb{R}^d,$$

where a_j, b_k are positive constants. By Proposition 6.5, $\mathcal{H}_{K_a} \lesssim \mathcal{H}_{K_b}$ if $N \leq M$. However, asking K_b to be a refinement kernel of K_a would impose a strong additional requirement that $a_j = b_j$ for all $j \in \mathbb{N}_N$. A more concrete example is the kernels K_p, K_q appeared in (4.7) and (4.8). By our discussion, if $p < q$ then K_q is a weak refinement but not a refinement of K_p .

References

- [1] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.
- [2] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Boston, MA, 2004.
- [3] S. Bochner, *Lectures on Fourier Integrals with an author's supplement on monotonic functions, Stieltjes integrals, and harmonic analysis*, Annals of Mathematics Studies **42**, Princeton University Press, New Jersey, 1959.
- [4] F. Cucker and S. Smale, On the mathematical foundations of learning, *Bull. Amer. Math. Soc.* **39** (2002), 1–49.
- [5] M. Cui and F. Geng, Solving singular two-point boundary value problem in reproducing kernel space, *J. Comput. Appl. Math.* **205** (2007), 6–15.
- [6] M. F. Driscoll, The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process, *Z. Wahrsch. Verw. Geb* **26** (1973), 309–316.
- [7] T. Evgeniou, M. Pontil and T. Poggio, Regularization networks and support vector machines, *Adv. Comput. Math.* **13** (2000), 1–50.
- [8] C. H. FitzGerald, C. A. Micchelli and A. Pinkus, Functions that preserve families of positive semidefinite matrices, *Linear Algebra Appl.* **221**, (1995), 83–102 .
- [9] K. Fukumizu, F. R. Bach and M. I. Jordan, Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces, *J. Mach. Learn. Res.* **5** (2004), 73–99.
- [10] C. Franke and R. Schaback, Solving partial differential equations by collocation using radial basis functions, *Appl. Math. Comput.* **93** (1998), 73–82.
- [11] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
- [12] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **209** (1909), 415–446.
- [13] M. Z. Nashed and G. G. Walter, General sampling theorems for functions in reproducing kernel Hilbert spaces, *Math. Control Signals Systems* **4** (1991), 363–390.
- [14] R. Opfer, Multiscale kernels, *Adv. Comput. Math.* **25** (2006), 357–380.
- [15] I. J. Schoenberg, Metric spaces and completely monotone functions, *Ann. of Math.(2)* **39**, (1938), 811–841 .
- [16] I. J. Schoenberg, Positive definite functions on spheres, *Duke. Math. J.* **9** (1942), 96–108.
- [17] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, 2002.
- [18] J. Shawe-Taylor, N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge, 2004.

- [19] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf and G. R.G. Lanckriet, Hilbert space embeddings and metrics on probability measures, *J. Mach. Learn. Res.* **11** (2010), 1517–1561.
- [20] E. M. Stein, *Singular Integrals And Differentiability Properties of Functions*, Princeton University Press, Princeton, 1971.
- [21] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [22] H. Wendland, *Scattered Data Approximation*, Cambridge University Press, Cambridge, 2005.
- [23] Y. Xu and H. Zhang, Refinable kernels, *J. Mach. Learn. Res.* **8** (2007), 2083–2120.
- [24] Y. Xu and H. Zhang, Refinement of reproducing kernels, *J. Mach. Learn. Res.* **10** (2009), 107–140.
- [25] N. D. Ylvisaker, On linear estimation for regression problems on time series, *Ann. Math. Statist.* **33** (1962), 1077–1084.
- [26] H. Zhang and J. Zhang, Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products, *Appl. Comput. Harmon. Anal.* **31** (2011), 1–25.